A Multi-Sensor Water Vapor Climate Data Record Using Cloud Classification

Eric J. Fetzer, Brian Wilson, Evan Fishbein, Brian H. Kahn, William G. Read, Joao Teixeira, Qing Yue, Van T. Dang and Alexandre Guillaume

Jet Propulsion Laboratory, California Institute of Technology

Introduction

We are creating a merged, multi-decadal climate data record of water vapor measurements from sensors on NASA's A-Train satellite constellation (see below for a list of acronyms). We are using this data record to characterize long-term changes in water vapor using observations with different noise characteristics, vertical resolution, and cloud-induced sampling. We are doing this by classifying individual water vapor sensors' scenes using collocated cloud observations. This permits the separation of water vapor variability from the effects of instrument noise, resolution and sampling. We will provide water vapor summaries within cloud classes from each of the water vapor sensors in the A-Train. Our work is being coordinated with the NVAP project (Forsythe et al., 2003) to exploit the highest quality observations from the A-Train first, and then extend backward in time into the multi-decadal record from operational sensors such as TOVS, SSM/I and ATOVS.

The data sources for our work (see Table 1) are large and diverse, necessitating new data technologies and analysis methods. The analyses often require data from separate satellite instruments whose observations are co-located and simultaneous. The co-registration of cloud and water vapor observations requires basic geometric information, but also familiarity with the observing properties of constituent data sets. A significant component of this work involves quantifying the relationship between

Instrument	Satellite	Product	Operational Period	Data Archive	Data Availability
EOS Data Sets					
CloudSat/ CALIPSO	CloudSat/ CALPIPSO	Clouds	June 2006-	CloudSat/ CALIPSO	On-line
MODIS	Aqua	Clouds	Aug 2002 –	LAADS	On-line
MLS	Aura	Upper tropospheric water vapor, cloud path	Sep 2004 –	JPL	On-line
AIRS	Aqua	Water vapor, temperature, cloud amount and height	Sep 2002 –	GDAAC	On-line
AMRS-E	Aqua	Total water vapor and cloud path	Jun 2002 –	RSS	On-line
AMSU MHS	POES Aqua METOP-A	Water vapor, temperature and cloud path	Jan 1999 –	CLASS	Partially on-line

Table 1. EOS era data sources. These data sets set have higher vertical range and resolution and are generally more stable than data from earlier instruments.

observations from different instruments. For example, Kahn et al. (2009) examined AIRS temperature and water vapor and CloudSat/CALIPSO clouds, and Fetzer et al. (2008) compared AIRS and MLS upper tropospheric water vapor. The constituent data sets used in this study are typically large in size: ~1-10 MB/day for records extending 5-10 years. Several data sets are being combined (AIRS, MLS, CloudSat, MODIS, AMSR-E) to relate water vapor information to co-located clouds.

Our Approach

The current standard practice in detecting climate trends is to search for long-term changes in a single geophysical parameter, focusing on the 'best' data from a given instrument. That practice can have significant shortcomings, since a detected trend can be caused either by a change in a geophysical quantity, or by shifts in cloud and precipitation states affecting viewing conditions. Our climatologies will be functions of scene typed by clouds, allowing the monitoring of water vapor changes within and between types. The basic tasks in this work are described in Table 2.

Table 2.
Project Goals:

- Create a multi-decadal earth science data record of water vapor from a variety of satellite sensors.
- Exploit collocated and coincident cloud observations to sort water vapor observations into climate types.
- Characterize the effects of clouds on water vapor observations.
- Provide a physically-based framework for weather and climate model comparisons.
- Examine trade-offs of instrument resolution, coverage, and record length.
- Use existing data sources and mature algorithms.

Implementation

Producing the proposed atmospheric water data records is a huge computational problem, which can only be made tractable by the use of modern Grid computing. Terabytes of data must be discovered and accessed from many data centers, and then sequenced through cloud-classification and statistical fusion operations, yielding terabytes of fused products. These products must be served to data analysts. Performing such large-scale computations is challenging, given finite disk space and compute resources, and repeating them in order to explore improved algorithms is daunting but necessary. To address the challenge we will leverage an existing distributed dataflow system, SciFlo. SciFlo nodes already exist within the AIRS and MISR areas at JPL and the NASA Goddard and Langley DAACs.

SciFlo (Wilson, et al., 2005; Yunck, et al., 2004) is a semantically-enabled ("smart") Grid Workflow system that ties together a peer-to-peer network of computers into an efficient engine for distributed computation. SciFlo leverages SOAP Web Services and the Grid Computing standards (WS-* & Globus Alliance toolkits), and enables scientists to do multi-instrument Earth Science by assembling reusable Web Services (SOAP or http GET one-line URL's), native executables, command-line scripts, and python codes into a distributed computing flow (operator graph).

Summary

The fundamental challenge in this work is creating integrated summaries of water vapor observations while also embodying information about associated clouds. While the observations used to create the summaries are spatially adjacent and simultaneous, these data sources are typically very large, heterogeneous and not designed for integrated analysis. Their complexity requires careful analyses by experienced scientists. Furthermore, managing the data sources in this research requires the application of advanced computational technology. These challenges must be addressed as we create a long-term record of the most important greenhouse gas.

Acronyms

A-Train	A flotilla of Earth-observing satellites composed of Aqua, Aura, CloudSat, PARASOL and CALIPSO satellites, flying in formation in low polar orbits.
AMSR-E	Advanced Microwave Scanning Radiometer – EOS (on Aqua)
AMSU	Advanced Microwave Sounding Unit (on Aqua and many NOAA sats.)
AIRS	Atmospheric Infrared Sounder (on Aqua)
ATOVS	Advanced TIROS Operational Vertical Sounder (on NOAA sats.)
CloudSat	Satellite flying a Millimeter-Wavelength Cloud Radar
DAAC	Distributed Active Archive Center
GDAAC	Goddard DAAC
JPL	Jet Propulsion Laboratory
LAADS	Level-1 and Atmosphere Archive and Distribution System
LDAAC	Langley DAAC (or LaRC)
METOP	European Operational Meteorology satellite
MISR	Multi-angle Imaging SpectroRadiometer
MLS	Microwave Limb Sounder the Aura satellite in the A-Train
MODIS	Moderate Resolution Imaging Spectroradiometer (on Terra & Aqua)
NVAP	NASA Water Vapor Project
OpenDAP	Open Data Access Protocol (a Data Grid technology)
RSS	Remote Sensing Systems (company & data archive)
SciFlo	A Scientific Dataflow (or Grid Workflow) Execution Engine
SOAP	Simple Object Access Protocol (a web services standard)
SSM/I	Special Sensor Microwave Imager (on DMSP sats.)
TOVS	TIROS Operational Vertical Sounder (on NOAA sats.)
UARS	Upper Atmosphere Research Satellite
WS	Web Services

References

Fetzer, E. J., W. G. Read, D. Waliser, B. H. Kahn, B. Tian, H. Vömel, F. W. Irion, H. Su, A. Eldering, M. de la Torre Juarez, J. Jiang and V. Dang (2008), Comparison of upper tropospheric water vapor observations from the Microwave Limb Sounder and Atmospheric Infrared Sounder, *J. Geophys. Res*., 113, D22110, doi:10.1029/2008JD010000.

Forsythe, J. M., D. L. Randel, S. Woo, D. S. McKague, and T. H. Vonder Haar, 2003: Extending the 12-year NVAP global water vapor dataset into the 21st century. Preprints, *12th AMS Conference on Satellite Meteorology and Oceanography*, Long Beach, California, Feb. 2003.

Kahn, B. H., A. Gettelman, E. J. Fetzer, A. Eldering, and C. K. Liang (2009), Cloudy and clear-sky relative humidity in the upper troposphere observed by the A-train, *J. Geophys. Res., 114*, D00H02, doi:10.1029/2009JD011738.

Wilson, B., B. Tang, G. Manipon, D. Mazzoni, E. Fetzer, A. Eldering, A. Braverman, E. Dobinson, and T. Yunck (2005), "GENESIS SciFlo: Scientific Knowledge Creation on the Grid Using a Semantically-Enabled Dataflow Execution Environment", peer-reviewed *Proceedings of SSDBM*.

Yunck, T., B. Wilson, A. Braverman, E. Dobinson and E. Fetzer (2004), "GENESIS: The General Earth Science Investigation Suite", *Proceedings of the Earth Science Technology Conference*, Palo Alto, CA.